Computational and Visual Support for Geographical Knowledge Construction: Filling in the Gaps Between *Exploration* and *Explanation*

Mark Gahegan and Boyan Brodaric[1]

GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, PA 16802. Tel: +1-814-865-2612, Fax: +1-814-863-7943, mng1@psu.edu, bmb184@psu.edu

Abstract

Although many different types of data mining tools have been developed for geographic analysis, the broader perspective of geographic knowledge discovery—the stages required and their computational support—have been largely overlooked. This paper describes the process of knowledge construction as a number of inter-related activities and the support of these activities in an integrated visual and computational environment, GeoVISTA *Studio*. Results are presented showing examples of each stage in the knowledge construction process and a summary of the inter-relationships between visualisation, computation, representation and reasoning is provided.
Keywords: knowledge discovery, data mining, visualisation, machine learning, abduction

1 Introduction

Despite enormous efforts in quantification, our understanding of many of the Earth's systems remains *non-axiomatic*; the systems are 'open' and consequently it is not possible to deduce all outcomes from known laws and rules. Geographic science must therefore adopt a manner that encourages the creation or uncovering of new knowledge (Baker, 1996; Takatsuka and Gahegan, 2001). For this reason alone—and completely uncoupled from concerns about increasing data volumes—it is vital that knowledge discovery methods can be brought successfully to bear on problems across geography and the wider geosciences.

[1] Also from Geological Survey of Canada.

The foundation of this paper is the relationship that knowledge construction and discovery activities have with the different approaches used for scientific inference; it is via an understanding of this relationship that we can categorise the kinds of knowledge that can be discovered or learned (epistemology), and thus begin to comprehend the roles required of domain experts and computational tools in the knowledge discovery process. Building on this foundation, the goal of the work described here is to facilitate the knowledge construction process in geography (or geo-sciences in general), by providing better computational support and closer integration of various exploratory visual and computational methods.

1.1 Background: Data Mining Knowledge Discovery to Date

Many different styles of spatio-temporal knowledge discovery have been proposed (see Roddick and Spiliopoulou, 1999 for a wide-ranging bibliography), from the entirely computational to visually-led methods (e.g. Openshaw *et al*., 1990; Koperski and Han, 1995; Knorr and Ng, 1996; Ester *et al*., 1996, 1998; Gahegan *et al*., 2001). Automated methods currently available concentrate on mining categories, clusters, outliers and other kinds of patterns that might occur in data, rather than on developing these patterns into knowledge structures. Examples are the Self Organising Map (Kohonen, 1997) and the AutoClass (Cheeseman and Stutz, 1996) and AutoClust (Estivill-Castro and Lee, 2001) packages.

Parallel efforts across many disciplines, including statistics, machine learning, databases and information visualisation have emphasized different aspects of the knowledge discovery process (e.g. Agrawal *et al*., 1993; Gehrke *et al*., 1999; Glymour *et al*., 1997; Rainsford and Roddick, 1999; Haslett *et al*., 1991; MacEachren *et al*., 1999). Within this spectrum, the various roles played by the expert and the machine differ greatly, sometimes with little thought to the different abilities that each has to offer (Valdez-Perez, 1999). The nature of the results obtained also differ, from probabilistic rules in computational form to deep insights gained by the expert from visual displays that cannot easily be shared or formalized. See Table 1 for a summary.

A specifically geographically-oriented overview of data mining and knowledge discovery has been recently developed as an emerging theme by the *University Consortium on Geographic Information Science* (UCGIS) (Yuan *et al*., 2001) and is available via http://www.ucgis.org/emerging/. The role played by geographic visualisation in supporting knowledge discovery activities is specifically tackled in a related research initiative by the *International Cartographic Association* (ICA) and reported by Gahegan *et al*. (2001). Section 3 below discusses the tasks and roles in the knowledge discovery process in more detail, but first we begin by examining some of the problems encountered by data mining and knowledge discovery methods applied in the geographical domain.

Table 1. Summary of the different discipline perspectives on data mining and knowledge discovery activities. Contributed by the author to Yuan *et al*. (2001).

	Databases	Statistics	Machine learning	Visualisation
Finding	Association rules	Local pattern analysis and global inferential tests	Neural networks, decision trees	Exploratory visualisation, Visual data mining
Reporting	Rule lists	Summary statistics, significance and power	Likelihood estimation, information gain	A stimulus within the visual domain
Representing	Schema update, metadata	Fitted statistical models, local or global	Rules, graphs, functions,	Shared between the scene and the observer
Validating	Weak significance testing	Statistical significance tests	Learning followed by verification	Human subjects testing.
Optimising	Reducing computationa l complexity	Data reduction and stratified sampling strategies	Stochastic search, gradient ascent methods	Hierarchical and adaptive methods, grand tours

2 Problems With Geographical Knowledge Discovery

2.1 What Constitutes Discovery?

Scientific discovery draws on a wide range of techniques, often simultaneously, in the search for new insights and theories (e.g. Hanson, 1958; Popper, 1959; Langley, 2000). Within this complex process, different forms of inference are required.

Philosophically, the process of actually discovering something new is closely tied to a form of reasoning called *abduction* (Psillos, 2000; Peirce, 1878). Abduction is the simultaneous act of uncovering some structure within the data and producing a hypothesis with which to explain it. The structure uncovered is necessarily unknown at the outset, while the theoretic explanation may be drawn from what is already known or may involve an expansion or a reshaping of existing knowledge. The importance of abductive reasoning in the geosciences is argued convincingly by Baker (1996).

Practically, Fayadd *et al*. (1996) also point to the development of knowledge via a number of stages: data selection, pre-processing, transformation, data mining and interpretation/evaluation that progressively refine a large dataset to the point where it makes sense to propose object structures and relationships.

2.2 Why is Geographical Knowledge Discovery Unique and Difficult?

Geography is an integrative discipline, so data necessarily spans a wide range of perspectives and interests, from the social to the physical and all points in between. Arising from this complex mix of perspectives, and coupled with a growing infrastructure for gathering information, the following problems arise.

1. **Data volume**. Like many disciplines where data mining is applied, geography is rich in data. Knowing which portions of a dataset to analyse, and which to ignore, becomes problematic.
2. **Complexities caused by data gathering and sampling**. Although data are available in increasing volume, it is still often the case that we must resort to surrogates or aggregates for the phenomena of interest, rather than direct measurements (Openshaw, 1984; Yuan *et al*., 2001).
3. **Complexities caused by local relationships**. Earth systems are so intrinsically interconnected that it is difficult to isolate an analysis conducted on some part of a system from the effects of other unmodelled aspects. The outcome often appears in statistical form as heteroskedasticity.
4. **Complexities associated with the domain itself**. Interesting and relevant signals in data are often entirely hidden by stronger patterns that must first be removed. For example, the cyclic nature of many geographical systems (daily, annual, sunspot) impose a heavy signal on data that can overshadow more localised variance (Roddick and Lees, 2001).
5. **Lack of appropriate methods.** While the existing techniques described in section 1.1 are useful for exploring a dataset, they fail to offer the explicit connection to theory or explanation that characterises abduction (section 2.1). The best that can be managed in a computational setting is a kind of low-level explanation offered in the language of the underlying feature-space, not in a higher form as domain knowledge.
6. **Difficulty in formalising the geographic domain**. There is, as yet, no universally accepted conceptual model of geography (e.g. Goodchild, 1992), and the models that are currently implemented in commercial GIS vary significantly one from another, often in quite fundamental, philosophical ways. This leads to three distinct problems: (**a**) data are often intrinsically non-commensurate, they cannot be directly compared or combined; (**b**) it is difficult to apply formal geographical knowledge to the process of knowledge discovery, since such knowledge is not readily available; (**c**) when new knowledge is uncovered it is difficult to represent that knowledge formally—there is nowhere to put it!

Additional details of some of these problems are presented by Yuan *et al.* (2001) and Miller and Han, (2001). All six of these problems speak to the need to bring domain knowledge to bear on the knowledge discovery process. As described above (point 6), the lack of a formal conceptual model for geographic information, models and processes, presents a formidable barrier to the automation of knowledge discovery and precludes the use of computationally-based abduction. As Psillos (2000) argues: "*The more conceptually adequate a model of abduction becomes, the less computationally tractable it is*".

In the absence of formal mechanisms for representing and applying domain knowledge, many researchers have modified the problem to focus on ways to engage the human as a direct node in the problem-solving process, rather than simply the consumer of the results. Using visually-led approaches, an abductive task is performed collaboratively between the observer and the visualisation. The stimulus to abduction—patterns in the visual displays—are observed as a consequence of the way the data are presented and the way the observer perceives and comprehends them. The simultaneous task of hypothesis generation is also similarly split, the mappings used to visualise the data may imply a hypothesis and an observer may generate one or more theories to explain the observed structure.

An additional problem is that, with few exceptions, the tools currently available operate in isolation, typically performing a single function (e.g. clustering, classification), or providing a view onto the data from a single perspective (e.g. scatterplot, parallel coordinate plot). By doing so, they implicitly assume that problems in science can be isolated to a single conceptual 'plane', which, when correctly understood and represented, can be fixed to form the basis on which further science can be constructed. However, it is often the case that we must experiment across more than one level simultaneously, particularly where complex situations and under-constrained theory present multiple alternatives that must be evaluated (Baker, 1999). Take, for example, the case of eco-regions: to be useful, an eco-region must not only neatly summarise many complex environmental dimensions, it must also form a useful basis for further analysis.

We thus regard *knowledge discovery* or *knowledge construction* within the geospatial sphere as a *developmental process*, with meaning being progressively constructed and refined through a series of pre-processing and interpretative steps (e.g. Fayyad *et al*., 1996; MacEachren, *et al*., 1999; Valdez-Perez, 1999; Risch *et al*., 1997; Ribarsky *et al*., 1999; Wong, 1999). Current systems lack adequate tools for supporting this process.

3 A Software Laboratory for Knowledge Construction

What we envisage is a computing environment where a user can move seamlessly between exploring data, constructing elements to represent observed structures, applying these structures operationally, assessing their performance and communicating findings. This has led us to construct GeoVISTA *Studio*, which aims to encompass this entire spectrum of activities in an integrated manner (Gahegan *et al*., 2002). A technical description of *Studio* has been previously reported (Takatsuka and Gahegan, 2001) so will not be repeated here. In short, it is a visual programming environment, which allows users to quickly design, test and refine strategies to explore and analyse geospatial data. Functionality is encapsulated in JavaBeans that support a range of activities, from visualising high dimensional feature spaces, applying neural networks and traditional statistical analysis tools through to mapping outcomes. Examples of the use of *Studio*, and its support for the process of discovery, appear in the following sub-section.

3.1. Scientific Activities in Support of Human-Directed Discovery

Although there is no consensus on any one scientific method, there are several prevalent activities proposed both by philosophers (e.g., Feyerband, 1975; Hanson, 1958; Kuhn, 1962; Popper, 1959) and geographers alike (e.g. Harvey, 1969). A subset of these activities is clearly pertinent to geographical knowledge discovery, beginning with exploratory activities from which concepts are synthesised then woven together into models or theories that can be evaluated and presented to others. Such activities are empirically supported by psychologists (Feist and Gorman, 1998; Zimmerman, 2000), and some are even implemented computationally (Langley, 2000; Shrager and Langley, 1990; Thagard, 1988), though largely without geographical focus.

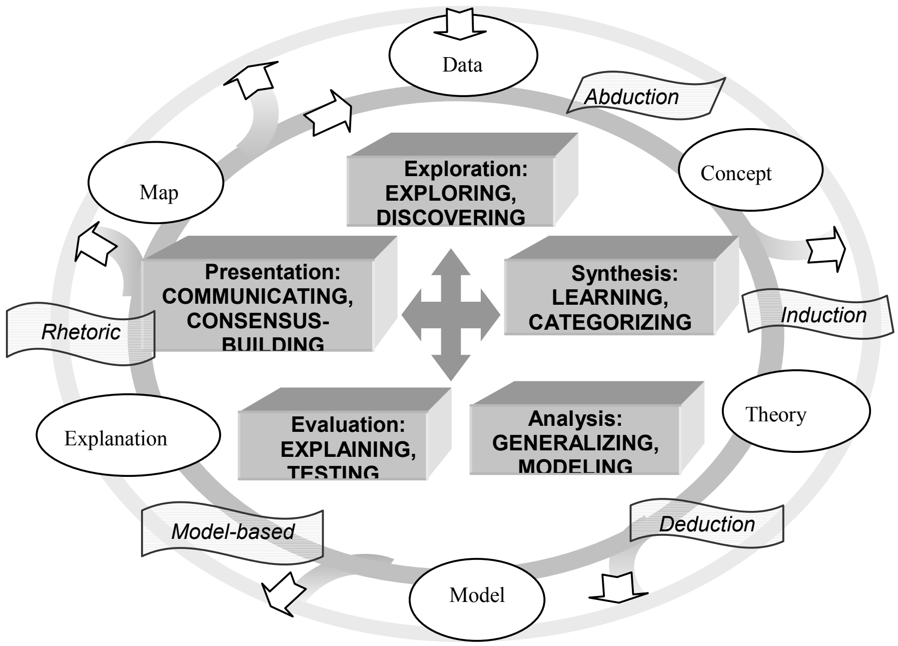

Fig. 1. A generalised framework for geoscientific discovery consisting of exploratory, synthesis, analysis, evaluation, and presentation activities. These respectively correlate with the evolution of features, concepts, theories and models, explanations, and presentations.

Fig. 1 presents an overview of some of the key aspects of the discovery process. It is depicted as a cycle, or a spiral, and this is apt because we can then envisage the negative outcomes causing the cycle to begin again and the positive outcomes (such as a validated theory described by a map) as being then becoming the data that is fed into the next iteration of the cycle at a more abstract level. Thus meaning is constructed in successive layers, each one supported by those below it (Popper, 1959). However, in practice, all stages are connected and may proceed in almost any order and simultaneously, as the connecting boxes in the interior of the diagram attempt to show. Further descriptions of the main stages follow.

Exploration: involves selecting which data, features (i.e. attributes), and feature weightings, are appropriate and significant, based on what is known and the prevailing scientific climate. It includes orienting human and other sensors to an environment by making preliminary observations. These observations can be notoriously subjective, being influenced by existing knowledge and various social pressures; moreover, this activity may be guided by perceptual and other implicit mechanisms that are difficult to express or even be conscious of (Shrager, 1990). Exploration is associated with scientific discovery inasmuch as it provides a stimulus for hypothesis generation and grounds for abduction; i.e. exploration involves selecting the data for which explanations are to be sought (abductively).

Fig. 2 shows two exploratory views provided by *Studio* onto a feature-space of state-level, socio-demographic data of the USA: a dynamic map and a clustering tool (a self-organising neural map or SOM). These and other tools provide the user with a number of different perspectives onto the data, ranging from the geographic distribution of variables (the map), to the clustering of places based on their similarity in feature space (SOM). The user can explore connections between these views via the linking and brushing tools (Cleveland and McGill, 1988; Buja *et al*., 1996; Hardisty *et al*., 2001) that *Studio* provides. A number of more traditional methods are also included here, such as the re-ordering of the feature space using correlation analysis and principal components. When used in conjunction with the visualisation tools, these allow the user to discard attributes that offer little or no additional information.

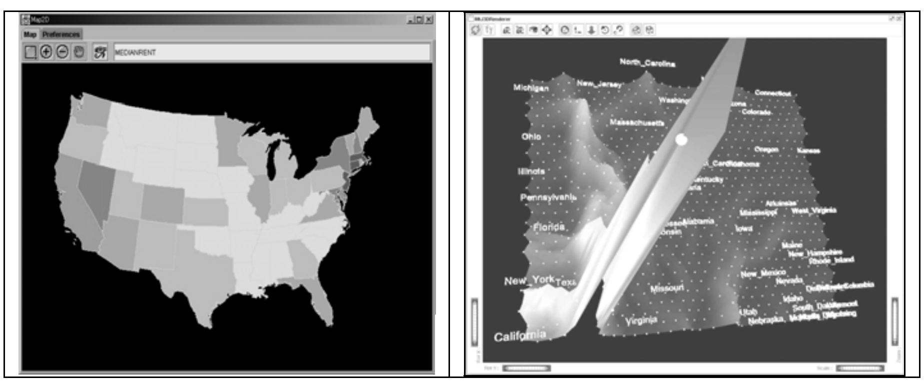

Fig. 2. Two views of a demographic dataset, allowing the user to explore the data from a number of different perspectives. Dynamic map (left), and a view of the unsupervised clustering of states provided by the SOM.

Synthesis: involves building concepts and taxonomies, or revising them, from selected features and data. These taxonomies then form the basis for terminology used in various explanatory structures, such as logical theories, mathematical laws or even text narratives. Synthesis is therefore closely aligned with the process of categorisation (in cognitive science) and classification (in computation), in which concepts (or classes) are explicitly induced from data. Induction is an important mechanism here, as classes are often determined by recognising recurring patterns in select feature dimensions of specific example data.

The following *Studio* examples (Fig. 3) show experiments in the development of categories for describing a complex forest habitat. The Parallel Coordinate Plot (PCP) is used, along with tools that allow user-driven groupings of the data to be defined and imposed. Categories are constructed visually by recolouring ranges of data across one or more dimensions, then synthesised inductively via a Learning Vector Quantisation (LVQ) classifier.

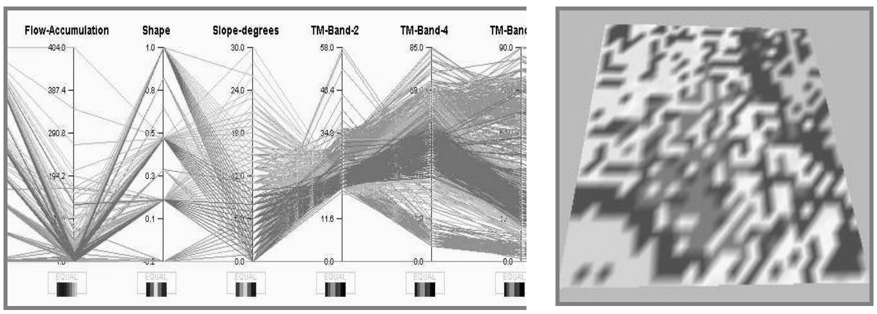

Fig. 3. Five candidate categories constructed in a PCP then learned using LVQ

Analysis: involves, on the one hand, using the established taxonomic framework and given data to develop general explanatory structures for concept behaviour and structure. On the other hand, it involves developing statements about how data are specifically related to each other (e.g. spatially, temporally, thematically, causally, etc.) and to the explanatory structures they exemplify. Taken together, these two aspects can be seen to form a model for the data; which can be either formal (e.g. probabilistic) or informal (e.g. text narrative, Suppes, 1960; 1962). Once a model is developed, the knowledge construction process in complete. Evaluation strategies must then establish its usefulness and reliability, and these strategies are often deduced from the model.

For example, the outcomes of synthesising various structures and categories (from the previous example) into themes can be used in the more traditional forms of GIS analysis (e.g. overlay) to build models of phenomena such as hydrology or landcover change (Fig. 4), and to suggest strategies for their evaluation.

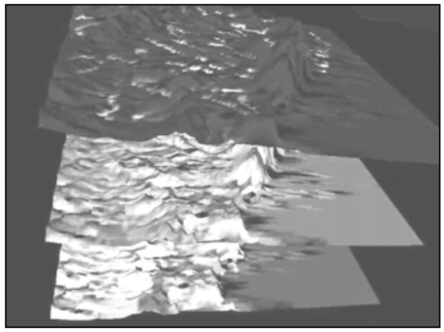

Fig. 4. Outcomes of synthesis form the inputs to analysis

Evaluation: involves testing the developed model against the validation data, or possibly against other models. It specifically requires data regularities to be explained by the model, using standards acceptable within a discipline. Philosophically and logically, unsuccessfully disconfirming the conceptual model is more informative than confirming it (Popper, 1959), a notion exploited by successful scientists who confirm early in their work, while model-building, but seek to disconfirm later as models evolve and settle (Feist & Gorman, 1998). The reasoning performed in evaluation is often model-based in that the behavioural, structural and logical-mathematical aspects of models are first proposed and then used as constraints on reasoning when testing models against the data.

Table 2 shows one form of evaluation where conflicts in meaning within a geological map are quantified; this evaluation can also be portrayed visually.

Table 2. Evaluation of semantic conflict between geologic mappers

Data Type	X	Y	r1	r2	Taxonomic Semantic Conflict (# sites)	Operational Semantic Conflict (# sites)	Total Semantic Conflict (% sites)	m Mean Conflict Distance	r Feature Space Radius	m/r
Structure	4	4	3	1	0	41	2.7 %	86.4	1635.45	0.0528
Lithology	7	7	5	2	2	155	10.61 %	158.38	40910.7	0.0038
Combined	10	10	8	2	1	137	10.43 %	133.50	40911.2	0.0036

Presentation: involves communicating with the external community to build consensus. This explicitly engages the social aspects of doing science (Kuhn, 1962), and possesses deep rhetorical components related to the holistic impact of narrative (Baker 1999; Ricoeur, 1985). Maps, and other visual devices in general, have always been an important knowledge construct in geography, and their rhetorical nature is well documented (e.g. Harley, 1989). Their transformation into digital products provides opportunity for elaborating communication methods between knowledge producers (e.g. map-based group collaboration and decision making systems (Jankowski *et al*., 2001)), and between producers and consumers (e.g. digital libraries, NRC, 1999).

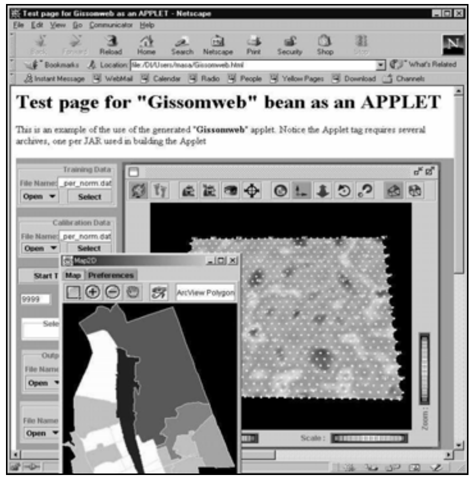

Fig. 5. Java Applet encapsulating the entire analysis of gentrification activities in Harrisburg, PA, that can be readily shared with others (Takatsuka and Gahegan, 2001)

The example in Fig. 5 shows a Java Applet, automatically created by *Studio* for web deployment, directly from an analysis of urban gentrification activities. It contains the entire process by which the analysis was constructed and is readily explorable by other researchers who can then accept or refute the findings.

4 Summary and Conclusions

The different knowledge construction activities are summarised in Table 3 according to the visualisation and computational techniques that can be used to support them via *Studio*. The dominant form of inference employed in each stage is also shown, as are the various representation devices required to support the activities. (The latter are not yet explicitly supported in *Studio* but are the subject of current and future work.)

Table 3. Knowledge discovery activities related to computational, visual, representation and reasoning issues

Scientific Activity	Visualisation	Computation	Representation		Reasoning
			Object	Structure	
Exploration	PCP, scatterplot, iconographic displays	SOM, k-means, clustering methods, GAM	Feature	Dataset	Abductive
Synthesis	Interactive visual classification, PCP	Machine learning, maximum likelihood, decision tree	Concept	Taxonomy	Inductive
Analysis	Scene composition, visual overlay	Statistical analysis	Rule Model	Theory	Deductive
Evaluation	Uncertainty visualisation	Statistical testing, M-C simulation	Inference	Explanation	Model-based
Presentation	Maps, charts, Reports, etc.	Web mapping, digital libraries, multi-media hypermaps, collaboratories	Document (e.g. Map)	Library (e.g. Atlas)	Rhetorical

Although presented in a coherent sequence above, these activities are generally thought to be mutually affective as depicted in Fig. 1. Of particular note is the fact that fundamental knowledge emergence is explicit within exploratory abductive activity. Fig. 1 also summarises the key representation and reasoning elements, which should act as a first order requirements statement for computational scientific systems. In addition to this, and following from the description in Sections 2 and 3, a geographical knowledge construction environment should:

1. Offer a variety of simultaneous views onto the data to gain alternative perspectives.
2. Allow different conceptual structures to be imposed onto the data (such as categories and relations) to prompt the generation of useful hypotheses.
3. Provide quick evaluation strategies for findings to assess their utility and viability in terms of existing theory, and means to revise proposed structures in the light of this evaluation, with little or no inertia.
4. Include higher-level analysis tools with which to formulate models based on knowledge construction outcomes, to test both the utility of the findings, and the correctness of the models themselves.
5. Incorporate tools to extract re-usable knowledge gained then represent it and communicate it to human experts (Gains, 1996).

6. Provide a packaging mechanism so that the entire knowledge discovery process can be shared and independently validated or refuted.

GeoVISTA *Studio* is able to touch on all of these issues to a limited extent at present, but at the time of writing concentrates on computational and visual support for the early discovery activities of exploration and synthesis. As such it can be seen as an environment within which to construct the categories and concepts that existing GIS make such heavy use of (but provide little support for their creation). *Studio* is available for no-cost download from http://www.geovistastudio.psu.edu/jsp/tryit.jsp, and we are actively seeking contributions of functionality (in the form of Java Beans) from other researchers.

Of the many challenges that remain, perhaps most difficult among them is to connect higher level, geographical domain knowledge with the tools used to search for potentially interesting anomalies or regularities within data. This higher level knowledge is required to support the abductive inference that converts data artifacts into useful domain knowledge (Sowa, 1999), and without it the structures uncovered are likely to remain unused. Taking a long-term perspective, it is unclear as to whether abduction might ultimately become fully automated, though such progress is highly unlikely in the short term. The current focus is thus to engage the head-knowledge of the domain expert as effectively as possible, using a variety of visual and computational tools brought to bear in a highly co-coordinated fashion.

Acknowledgements

The support of NSF for this work (under Digital Government grant EIA-9983445) is gratefully acknowledged.

References

Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. ACM SIGMOD, pp 207-216

Baker V (1996) Hypotheses and geomorphological reasoning. In: Rhoads, BL and Thorn, CE (eds) The scientific nature of geomorphology. Wiley, New York, pp 57-86

Baker V (1999) Geosemiosis. Geological Society of America Bulletin 111(5): 633-645

Brodaric B, Gahegan M (2001) Learning Geoscience Categories In Situ: Implications for Geographic Knowledge Representation. In: Proceedings, ACM-GIS 2001: The Ninth Association for Computing Machinery International Symposium on Advances in Geographic Information Systems. Atlanta, GA, Nov. 9-10, 2001

Buja A, Cook D, Swayne D (1996) Interactive high-dimensional data visualization. Journal of Computational and Graphical Statistics 5(1):78-99

Cheeseman P, Stutz J (1996) Bayesian Classification: Theory and results. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in Knowledge Discovery and Data Mining, Cambridge, MA: AAAI/MIT Press, pp 153-189

Cleveland WS, McGill ME (1988) Dynamic Graphics for Statistics. Wadsworth & Brookes/Cole, Belmont, California, USA

Ester M, Kriegel H-P, Sander J, Xu X (1996) A Density Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). pp 226-231

Ester M, Kriegel H-P, Sander J (1998) Algorithms for characterization and trend detection in spatial databases. In: Proc. 4th International Conference on Knowledge Discovery and Data Mining (KDD'98). New York, USA, pp 44-50

Estivill-Castro V, Lee I (2001) Argument free clustering for large spatial point-data sets via boundary extraction from Delaunay diagram. Computers, Environment and Urban Systems (in press)

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Magazine, Fall 1996, pp 37-54

Feist GJ, Gorman ME (1998) The psychology of science: review and integration of a nascent discipline. Review of general psychology 2(1): 3-47

Feyerband P (1975) Against Method. London: Verso

Gahegan M, Wachowicz M, Harrower M, Rhyne TM (2001) The integration of geographic visualization with knowledge discovery in databases and geocomputation. Cartography and Geographic Information Systems (special issue on the ICA research agenda) 28(1): 29-44

Gahegan M, Takatsuka M, Wheeler M, Hardisty F (2002) GeoVISTA Studio: a geocomputational workbench. Computers, Environment and Urban Systems (in press)

Gains BR (1996) Transforming Rules and Trees into Comprehensible Knowledge Structures. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, Cambridge, MA

Gehrke J, Ganti V, Ramrakrishnan R, Loh W-Y (1999) BOAT—Optimistic decision tree construction. In: Proc. SIGMOD 1999, ACM Press: New York, pp 169-180

Glymour C, Madigan D, Pregibon D, Smyth P (1997) Statistical themes and lessons for data mining. Journal of Data Mining and Knowledge Discovery 1:11-28

Goodchild MF (1992) Geographical data modeling. Computers and Geosciences 18(4): 401-408

Hanson N (1958) Patterns of discovery. Cambridge University Press, Cambridge

Hardisty F, MacEachren AM, Gahegan M, Takatsuka M, Wheeler M (2001) Cartographic Animation in Three Dimensions: Experimenting with the Scene Graph. In: 20th International Cartographic Conference (Beijing, China, August 6-10, 2001)

Harley JB (1989) Deconstructing the map. Cartographica 26(2): 1-20

Haslett J, Bradley R, Craig P, Unwin A, Wills G (1991) Dynamic graphics for exploring spatial data with application to locating global and local anomalies. The American Statistician 45(3): 234-242

Jankowski P, Andrienko N, Andrienko G (2001) Map-centred exploratory approach to multiple criteria spatial decision making. International Journal of Geographical Science 15(2): 101-127

Kohonen T (1997) Self-organizing maps. Berlin, New York

Knorr EM, Ng RT (1996) Finding aggregate proximity relationships and commonalities in spatial data mining. IEEE Transactions on Knowledge and Data Engineering 8(6): 884-897

Koperski K, Han J (1995) Discovery of Spatial Association Rules in Geographic Information Databases. In: Proc. 4th International Symposium on Large Spatial Databases (SSD95). Maine, pp 47-66

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Langley P (2000) The computational support of scientific discovery. Int. Journal of Human-Computer Studies 53: 393-410

MacEachren AM, Wachowicz M, Edsall R, Haug D, Masters R (1999) Constructing knowledge from multivariate spatio-temporal data: integrating geographical visualization with knowledge discovery in database methods. International Journal of Geographic Information Science 13(4): 311-334

Miller HJ, Han J (2001) Geographic data mining and knowledge discovery: An overview. In: Miller HJ and Han J (eds) Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, pp 3-32

NRC (1999) National Research Council Distributed Geolibraries: Spatial Information Resources, Summary of Workshop, Washington, DC: National Academy Press

Openshaw S (1984) The modifiable areal unit problem. CATMOG 38, (Norwich: Geo Abstracts)

Openshaw S, Cross A, Charlton M (1990) Building a Prototype Geographical Correlates Machine. International Journal of Geographical Information Systems 4(4): 297-312

Peirce CS (1878) Deduction, induction and hypothesis. Popular Science Monthly 13: 470-482

Popper K (1959) The logic of scientific discovery. Basic Books: New York

Psillos S (2000) Abduction: Between conceptual richness and computational complexity. In: Flach PA, Kakas AC (eds) Abduction and Induction. Dordrecht: Kluwer, pp 59-74

Rainsford CP, Roddick JF (1999) Database issues in knowledge discovery and data mining. Australian Journal of Information Systems 6(2): 101-128

Ribarsky W, Katz J, Holland A (1999) Discovery visualization using fast clustering. IEEE Computer Graphics and Applications. September/October 1999, pp 32-39

Ricoeur P (1985) Time and Narrative. McLaughlin K, Pellauer D (trans.), University of Chicago Press, Chicago

Risch JS, Rex DB, Dawson ST, Walters TB, May RA, Moon BD (1997) The STARLIGHT information visualization system. In: IEEE Proceedings, International Conference on Information Visualization '97. pp 42-49

Roddick JF, Lees B (2001) Paradigms for spatial and spatio-temporal data mining. In: Miller HJ, Han J (eds) Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, pp 33-50

Roddick JF, Spiliopoulou M (1999) A bibliography of temporal, spatial and spatio-temporal data mining research [online]. SIGKDD Explorations. 1(1) Available from: http://www.cis.unisa.edu.au/~cisjfr/STDMPapers/

Shrager J (1990) Commonsense Perception and the psychology of theory formation. In: Shrager J, Langley P (eds.) Computational Models of Scientific Discovery and Theory Formation. Morgan Kaufman, San Mateo, pp 437-470

Shrager J, Langley P (1990) (eds) Computational Models of Scientific Discovery and Theory Formation. Morgan Kaufman, San Mateo

Sowa JF (1999) Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole, Pacific Grove, CA

Suppes P (1960) A comparison of the meaning and uses of models in mathematics and the empirical sciences. In: Suppes P (ed) Studies in the Methodology and Foundations of Science. Reidel, Dordrecht

Suppes P (1962) Models of Data. In: Nagel E, Suppes P and Tarski A (eds) Logic, Methodology and the Philosophy of Science: proceedings of the 1960 International Congress. Stanford University Press, Stanford, CA, pp 252-61

Takatsuka M, Gahegan M (2001) GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. Computers and Geosciences (In Press)

Thagard P (1988) Computational philosophy of science. MIT Press, Cambridge, Mass

Valdez-Perez RE (1999) Principles of human computer collaboration for knowledge discovery in science. Artificial Intelligence 107(2): 335-346

Wong PC (1999) Visual data mining. IEEE Computer Graphics and Applications 19(5): 20-21

Yuan M, Buttenfield B, Gahegan M, Miller H (2001) Geospatial Data Mining and Knowledge Discovery [online]. A UCGIS White Paper on Emergent Research Themes. Available from: http://www.ucgis.org/emerging/

Zimmerman C (2000) The development of scientific reasoning skills. Developmental Review 20: 99-149